

Statistics 253/352  
Spatial Statistics  
Assignment #2  
Due Tuesday, June 8, 2009

Prof. J. Taylor

- Q. 1) This question introduces the ICM (Iterated Conditional Modes) algorithm [Bes86], an algorithm used often for image segmentation, which we will compare with the usual EM algorithm for fitting finite mixture models. The data for this question is a  $109 \times 91$  2d slice of an anatomical MRI

<http://www-stat.stanford.edu/~jtaylor/courses/stats352/slice.csv>

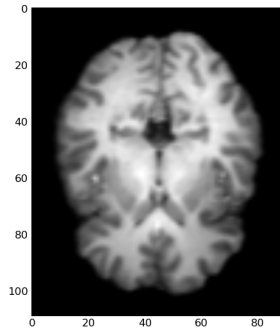
and the goal is to segment, or partition it into several regions: white matter, grey matter, CSF (cerebro-spinal fluid).

A description of the EM algorithm for two classes can be found in several places, e.g. Ch. 8 of [HTF01]. The multi-class version goes as follows.

**Algorithm 1 Initialization** Choose initial estimates  $(\hat{\mu}_k^{(0)}, \hat{\sigma}_k^{(0)}, \hat{\pi}_k^{(t)})_{1 \leq k \leq K}$ .

**E-step** Having reached  $t$  iterations, compute “responsibilities”

$$\hat{\gamma}_{ik}^{(t+1)} = \frac{\hat{\pi}_k^{(t)} f(Y_i; \hat{\mu}_k^{(t)}, \hat{\sigma}_k^{(t)})}{\sum_{l=1}^K \hat{\pi}_l^{(t)} f(Y_i; \hat{\mu}_l^{(t)}, \hat{\sigma}_l^{(t)})}$$



where  $f(y; \mu, \sigma) = e^{-(y-\mu)^2/2\sigma^2} \sqrt{2\pi\sigma^2}$  is the univariate Gaussian density.

**M-step** Re-estimate the class mean and variances

$$\begin{aligned}\hat{\mu}_k^{(t+1)} &= \frac{\sum_{i=1}^n \hat{\gamma}_{ik}^{(t+1)} Y_i}{\sum_{i=1}^n \hat{\gamma}_{ik}^{(t+1)}} \\ \hat{\sigma}_k^2{}^{(t+1)} &= \frac{\sum_{i=1}^n \hat{\gamma}_{ik}^{(t+1)} (Y_i - \hat{\mu}_k^{(t+1)})^2}{\sum_{i=1}^n \hat{\gamma}_{ik}^{(t+1)}}\end{aligned}$$

**Labelling** Repeat the E and M-steps until convergence and assign final class labels  $\hat{L}_i = \operatorname{argmax}_{1 \leq k \leq K} \gamma_{ik}$ .

The ICM algorithm is very similar, but introduces spatial dependence in the labels in the labels. Formally, the finite mixture model assumes labels in  $\{1, \dots, K\}$  are assigned independently to pixels according to some multinomial with probabilities  $(\pi_1, \dots, \pi_K)$  and within each class the observed data is  $N(\mu_k, \sigma_k^2)$ . The EM algorithm tries to estimate  $(\pi_l, \mu_k, \sigma_k^2, 1 \leq k \leq K)$  as well as the posterior probabilities

$$\mathbb{P}(L_i = k | Y_i = y_i) = \frac{\pi_k f(y_i; \mu_k, \sigma_k)}{\sum_{l=1}^K \pi_l f(y_i; \mu_l, \sigma_l)}$$

which are estimated by the ‘‘responsibilities’’ above, and classification is done by assigning pixels to the highest (estimated) posterior probability.

For the ICM model, independent label assignment is replaced with a Markov random field  $L_i$  with labels generated from a generalization of the Ising model called the Potts model. The potential for a configuration of labels  $L$  in the Potts model with no external field is

$$U(L) = \beta \sum_{i \sim j} \delta_{L_i, L_j}$$

but the field with an external potential is more easily described in terms of indicators  $I_{ik}, k = 1, \dots, K, i = 1, \dots, N = 109 * 91$  defined by

$$I_{ik}(L) = \begin{cases} 1 & L_i = k \\ 0 & \text{otherwise.} \end{cases}$$

In this form, the potential, with external fields  $\alpha$  is

$$U(L) = \sum_{k=1}^K \alpha_{ik} I_{ik}(L) + \beta \sum_{i \sim j} \sum_{k=1}^K I_{ik}(L) I_{jk}(L). \quad (1)$$

Finally, the ‘‘iterated’’ in ICM relates to relabelling the pixels in every iteration of the E-M steps.

**Algorithm 2 Initialization** Choose initial estimates  $(\hat{\mu}_k^{(0)}, \hat{\sigma}_k^{(0)}, 1 \leq k \leq K)$  and an initial configuration of labels  $(L_i^{(0)}, i = 1, \dots, N)$ .

**Estimation** Given the  $t$ -th configuration of labels  $L^{(t)} = (L_i^{(t)}, i = 1, \dots, N)$ , estimate the parameters for the Normal model

$$\hat{\mu}_k^{(t+1)} = \frac{\sum_{i=1}^N I_{ik}(L^{(t)})Y_i}{\sum_{i=1}^N I_{ik}(L^{(t)})}$$

$$\hat{\sigma}_k^{(t+1)} = \frac{\sum_{i=1}^N I_{ik}(L^{(t)})(Y_i - \hat{\mu}_k^{(t+1)})^2}{\sum_{i=1}^N I_{ik}(L^{(t)})}$$

**Labelling** For  $i = 1, \dots, N$  compute the analogous (conditional) “responsibilities”

$$\hat{\gamma}_{ik}^{(t+1)} = P\left(L_i = k \mid L_j = L_j^{(t)}, j \neq i, Y_i; \hat{\mu}_k^{(t)}, \hat{\sigma}_k^{(t)}, 1 \leq k \leq K\right)$$

and assign labels

$$L_i^{(t+1)} = \underset{1 \leq k \leq K}{\operatorname{argmax}} \hat{\gamma}_{ik}^{(t+1)}.$$

**Fitting** Repeat **Estimation** and **Labelling** step until convergence.

- (a) Plot a histogram of the image intensities. Does this histogram seem appropriate for a mixture model?
- (b) Fit a finite mixture model to the image data  $Y_i, i = 1, \dots, N$  with Gaussian components using the EM algorithm with  $K = 5$  to 10 components, starting with initial values derived from the quantiles of the histogram (several components will be taken by the many zeros in the image).  
Plot the resulting classification using, for instance `imshow` in R. Plot the fitted density along with the histogram. How well does the model fit? (Note: to avoid a problem with many 0’s in the image, you may have to enforce some arbitrary lower bound for the estimates of  $\sigma$  within each class).
- (c) The finite mixture model setting, i.e. with labels assigned independently, can be thought of as a special case of the setup in the ICM model in which the second term in (1) is 0. Give suitable choices for the fields  $\alpha_{ik}$  so that this potential corresponds to the posterior distribution of the labels  $(L_i, 1 \leq i \leq N)$  given the data  $(Y_i, 1 \leq i \leq N)$ .
- (d) Argue that

$$L_i \mid L_j, j \neq i, Y \sim \text{Multinomial}(1, p(L_j, j \neq i, Y)).$$

Give a formula for  $p$ , up to a proportionality constant, for a given choice of  $\beta$  and fields  $\alpha_{ik}$ . Fit the ICM model with the same fields as in (c) for several choices of the parameter  $\beta$ . How do the results of the EM and the ICM model compare for this sample?

- (e) Propose a “softer” version of the ICM using the responsibilities  $\hat{\gamma}_{ik}^{(t)}$ . Does this substantially change the fit?
- (f) Try adding independent  $N(0, \sigma^2)$  noise to the image. How large do you have to make  $\sigma$  to see a noticeable difference between the finite mixture model EM and the ICM algorithms?

Q. 2) This question uses some of the disease mapping tools to model incidences of various diseases in California by county. I managed to find a map of the California counties from

<http://casil.ucdavis.edu/casil/>

This map can be loaded into R using this little script

<http://www-stat.stanford.edu/~jtaylo/courses/stats352/R/loadCA.R>

which assumes you’ve downloaded

<http://www-stat.stanford.edu/~jtaylo/courses/stats352/data/nty24k97.shp>

and will store the results in an object known as `ca.poly`. This object is much like the `nc` object in the disease mapping examples.

(a) The file <http://www-stat.stanford.edu/~jtaylo/courses/stats352/data/ca`data.csv>

contains the incidences of diseases

- confirmed H1N1 diagnoses (`confirmed`);
- probable H1N1 diagnoses (`probable`);
- HIV (this number seems quite low because of an apparent change in the way HIV/AIDS were counted in 2006, and only seems to reflect incidence since 2006);
- AIDS (this number was the reporting standard up to 2006);
- West Nile virus, a mosquito-borne illness (these are the total number of incidents per county in 2007 and 2008, and include human cases, animal cases as well as positive tests of mosquitoes).

The results are by county in California. Also included are the names of the county seats in each county, their population and, finally, their approximate latitude and longitude. All the data were collected from

<http://ww2.cdph.ca.gov/Pages/default.aspx>

Q. 3) Pooling H1N1 as confirmed + observed, and HIV as HIV+AIDS, plot the raw incidence of each disease on a map of the counties in California using the function `sppplot`. (Note, due to some disconnected components of counties, there are more polygons in the data than counties so you must assign the same value to each polygon in the county to get a proper map). Do any of the disease incidences show any distinguishing spatial features?

Q. 4) Plot the analogous “standardized mortality ratio” for each of H1N1, HIV and West Nile. Do any of the disease rates show any distinguishing spatial features?

Q. 5) Using the `DCluster` package as in the R example

<http://www-stat.stanford.edu/~jtaylo/courses/stats352/R/disease`map.R>

fit the Poisson-Gamma and lognormal empirical Bayes disease mapping model of Clayton and Kaldor and replot the smoothed rates. Do you notice significant differences from (b)?

Q. 6) The table

<http://www-stat.stanford.edu/~jtaylo/courses/stats352/R/ca`distances.table>

has the distances between each of the county seats in California. The columns and rows are ordered identically, so, for instance, the 5th row of numbers corresponds to distances from San Andreas, the seat of Calaveras county, to each other county seat in the state. Using various radii, construct a graph of the counties based on whether or not the distance is less than a given radius. Fit an auto-Poisson model using

$$\theta_{ij} = \begin{cases} \gamma d_{ij} & i \sim j \\ 0 & \text{otherwise.} \end{cases}$$

Plot the image of conditional Poisson means for each county. Are all the estimated values of  $\gamma$  negative? Do you expect them to be?

Q. 7) **Bonus:** If you have extra energy, try fitting Besag’s lognormal model described in [BYM91].

## References

- [Bes86] Julian Besag. On the statistical analysis of dirty pictures. *J. Roy. Statist. Soc. Ser. B*, 48(3):259–302, 1986.
- [BYM91] Julian Besag, Jeremy York, and Annie Mollié. Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Statist. Math.*, 43(1):1–59, 1991. With discussion and a reply by Besag.
- [HTF01] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning; Data Mining, Inference and Prediction*. Springer Verlag, New York, 2001.